Crypto Data and Finance

Igor Makarov
LSE

Columbia, October 21, 2022

Agenda

- Crypto data
- Blockchain analysis of the Bitcoin market

Crypto data

- Off-chain: Not recorded on the blockchain
  - CEX and other centralized intermediaries
- On-chain: Recorded on the blockchain
  - Coin transfers
  - dApps (smart contracts automating financial services)

Off-chain data

- Crypto CEX operate similar to traditional exchanges
  - Limit order book
  - Spot and derivatives contracts
  - Investor level data are proprietary
- Current book level and transaction data are typically freely available via exchange APIs
- Historical data
  - Exchange-on-exchange basis
  - Public web-sites, e.g., `http://api.bitcoincharts.com/v1/csv/`
  - Specialized data providers, e.g., Kaiko
- Exchanges are non-integrated: A unique laboratory for studying arbitrage and price formation, see e.g., Makarov and Schoar (2020)

On-chain data

- Blockchain technology introduces new ways of storing data and new set of intermediaries
- Data
  - Transactions are public
  - Pseudonymous cryptocurrency addresses
- New intermediaries
  - Validators (miners)
  - dApps

Data challenges

- Crypto addresses are easy to generate ⇒ potentially many addresses belong to the same entity
- Link anonymous addresses to real-life entities
- Processing smart contract data
- Cross-chain analysis

Blockchain analysis of the Bitcoin market

- Systematic analysis of the Bitcoin market using blockchain data
  - We build a novel Bitcoin database and methodology for identifying information about the main market participants
- Three major pieces of analysis:
  - **Network structure:** Analyze the transaction volume and network structure of the main market participants
    - Exchanges are central entities; 75% of volume is linked to exchanges
  - **Ownership concentration:** Document the ownership concentration of the largest bitcoin investors
    - High concentration: top 0.01% investors control 24% of bitcoins in circulation
  - **Miners:** Study the concentration and regional composition of miners who ensure the integrity of the Bitcoin blockchain
    - High concentration: 50 miners often control 50% of the total market

Data construction (1)

- Use the BlockSci program to convert raw data into a database
- Clustering: The Bitcoin community developed clustering heuristics to assign addresses to the same entity. We start with the most conservative Union-Find algorithm
  - All transactions that share inputs belong to the same entity

  - Lower bound on the size of individual entities, since with a bit of effort a user can ensure any address is used only once and not mixed with other user's addresses in the same transaction

Data construction (2)

- Link addresses to real-life entities using public and proprietary sources
  - Scrape cryptocurrency blogs and websites, such as Reddit, Blockchain.info, bitcointalk.org, etc
  - State-of-the-art database of crypto entities from Bitfury Crystal Blockchain, one of the leading providers of anti-money-laundering tools
- Most complete information about crypto entities that have been used in academic research
  - We cover 3,500+ of the largest entities
  - 700+ exchanges and OTC desks, 113 gambling sites, 57 on-line wallets, 120+ mining pools and miners, 400+ scammers, 1700+ dark net marketplaces and illegal services

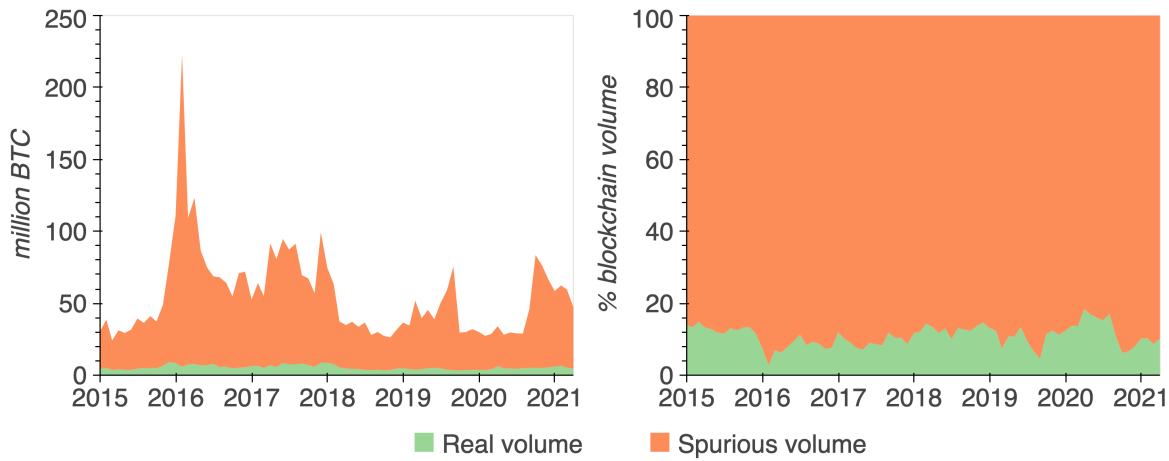
Data construction (3)

- We develop methodology to obtain information about other important Bitcoin entities:
  - To make sure that we are not missing major players on the blockchain, we analyze the top 10,000 unknown clusters with the largest Bitcoin volume
    - We use the observed transaction patterns of known exchanges and OTC brokers to infer which unidentified clusters are likely exchanges or OTC brokers
  - To analyze ownership concentration we need to differentiate between addresses belonging to individual investors and those belonging to intermediaries
    - We use graph analysis and examine utilization pattern to separate intermediary and individual accounts
  - We identify miners by tracing rewards distribution of the largest mining pools to individual miners

Transaction Volume and Network Structure

Spurious Volume

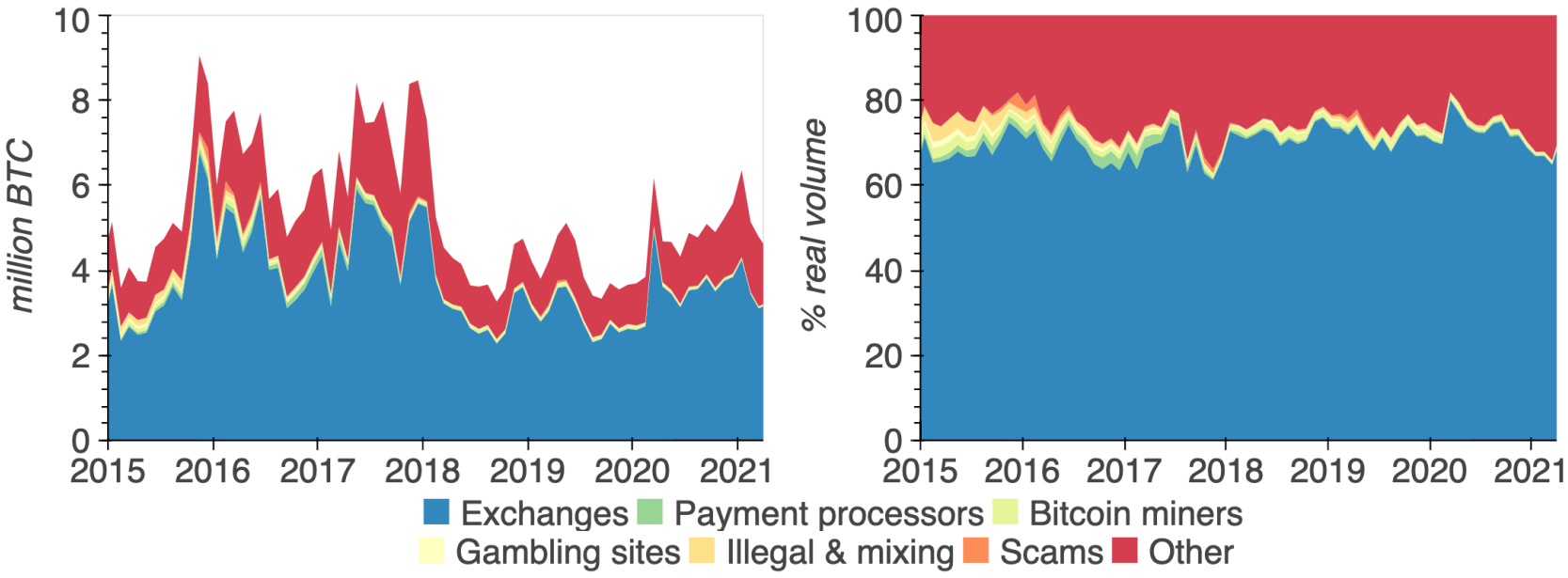
- The UTXO design and the preference of many of users for anonymity lead to spurious volume: transactions where an address sends its balance to itself or to another address controlled by the same entity



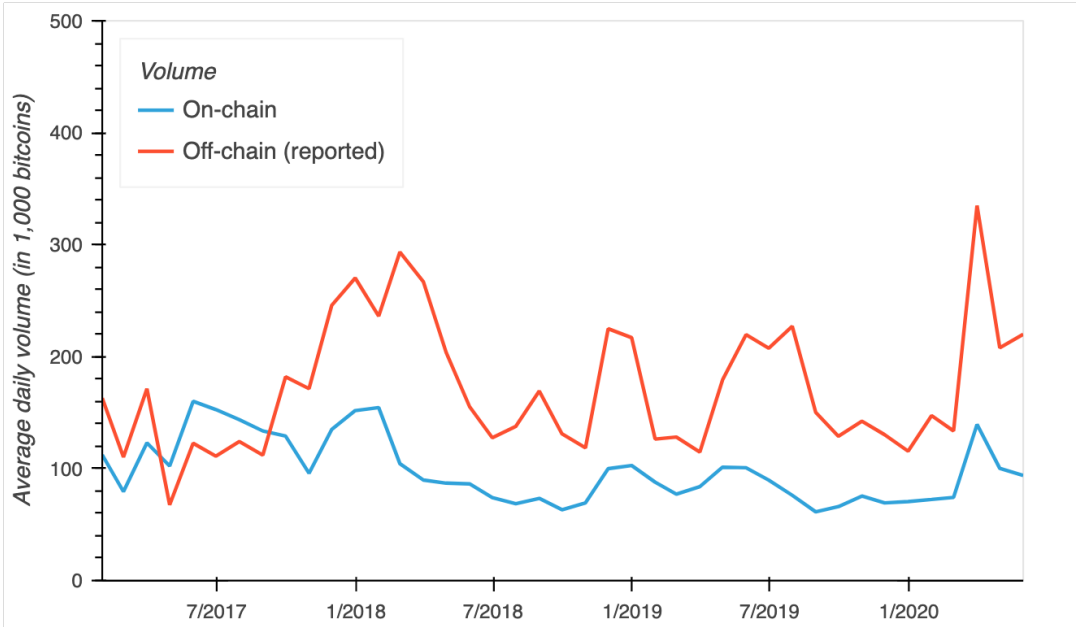
- The number of new addresses or nominal volume might not be meaningful statistics

Real Volume Decomposition

- Majority of volume on the Blockchain is for trading activity



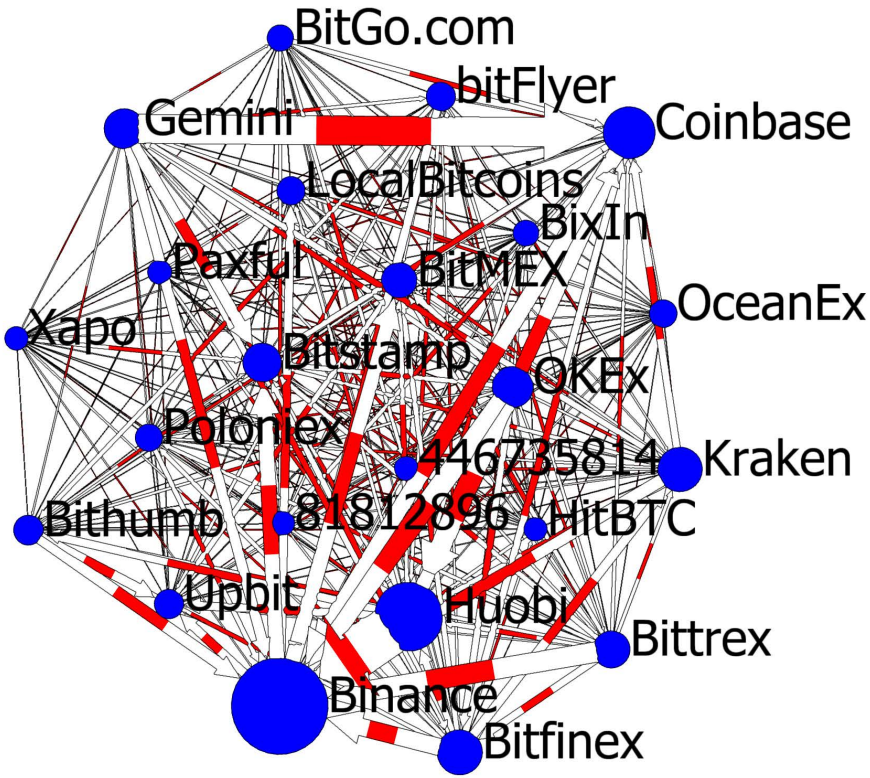
Off-chain and one-chain exchange volume



- Trading volume of BTC against the four major fiat currencies and two stablecoins: USD, EUR, JPY, KRW, USDT, and USDC
- The weekly correlation is 44%

Bitcoin Network

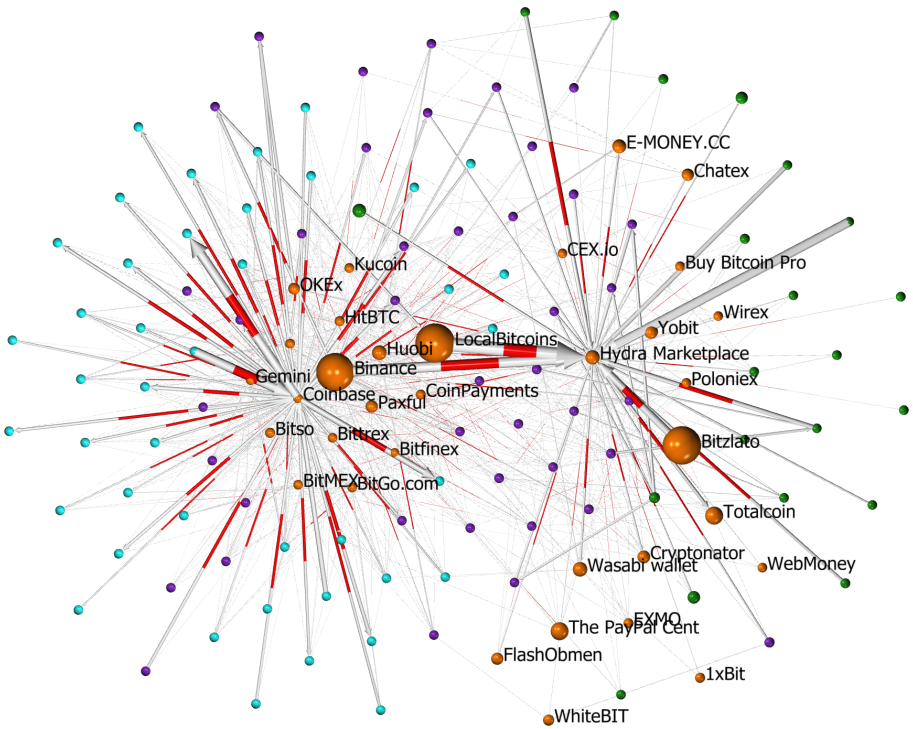
- Exchanges are central entities and are highly interconnected



- Ex: Network of entities that receive > 500K BTC over 2018-2020
  - 18 exchanges, 3 online wallets, 2 unknown entities — likely large OTC trading desks
  - Almost a complete graph

Illegal Transactions (Hydra Market)

- A small share of total volume (< 3%) but not a trivial amount ~ $2.4B in 2020



Ex: Hydra network: Retain only nodes that send >1000 BTC within the network

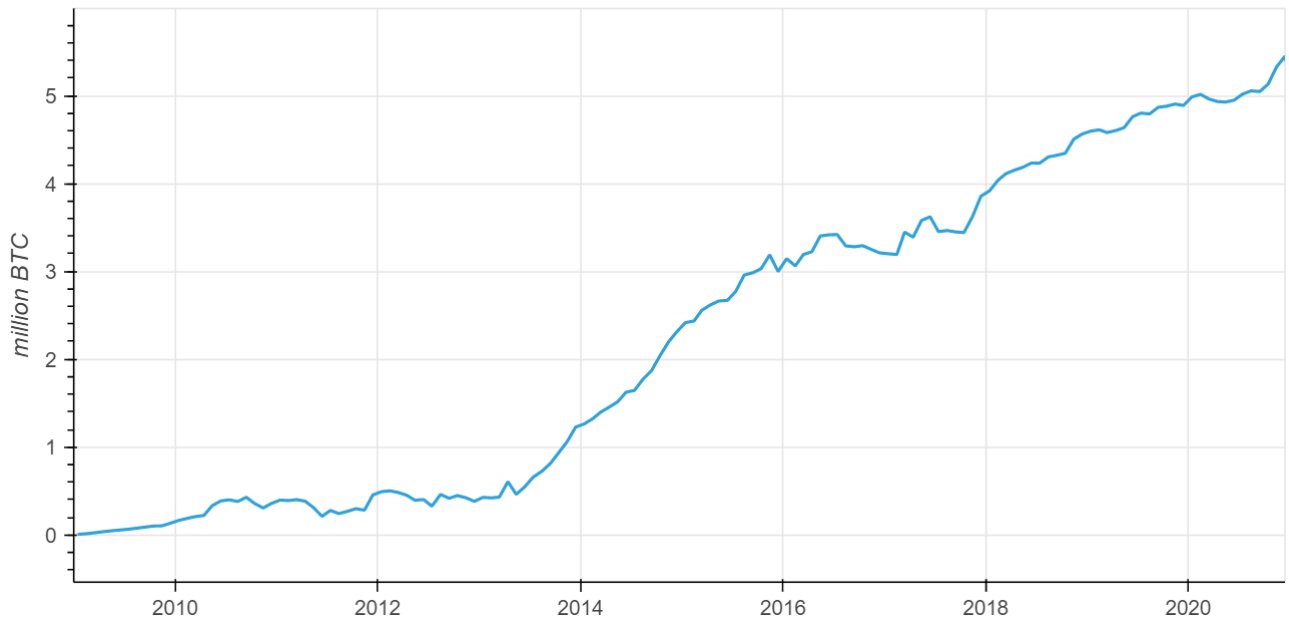
- Exchanges with lax KYC norms with serve as a gateway for money laundering and other gray activities
  - E.g. LocalBitcoins, Bitzlato, Binance
  - Once the flows arrive at these exchanges they get mixed with other flows and become virtually untraceable, and so can be sent anywhere afterwards

Ownership of Bitcoin

Ownership of Bitcoin

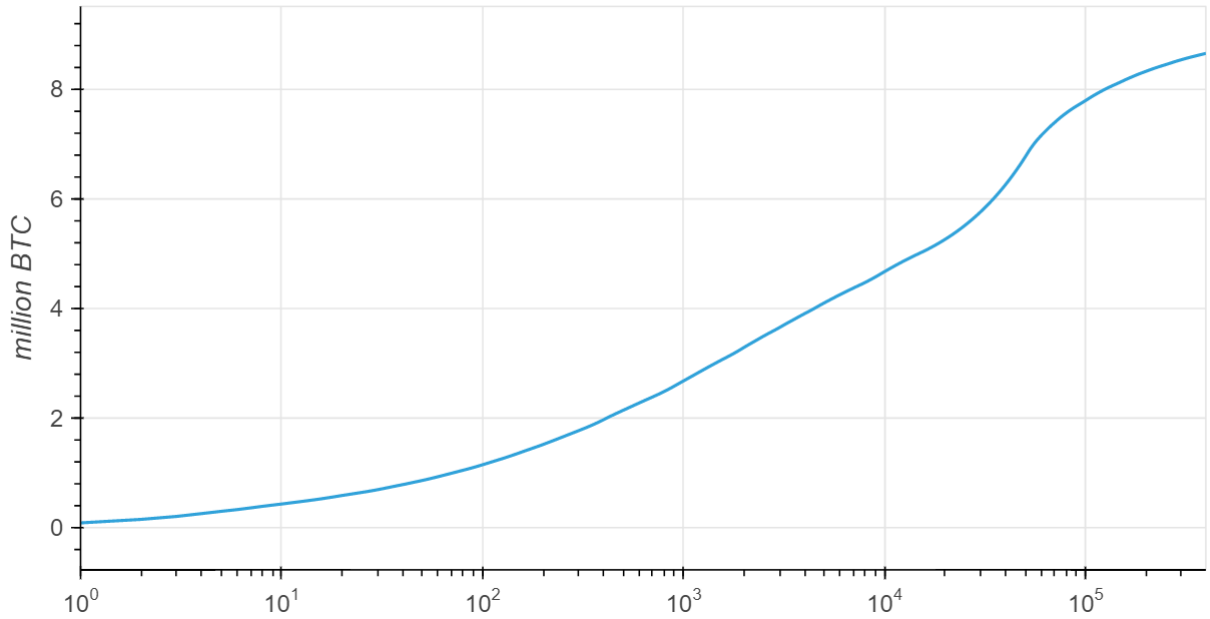
- Important to understand ownership and concentration of Bitcoin holdings
  - Determines who will benefit most from wider adoption. A select few investors or the general public?
- A challenging task:
  - More than just tracing “rich list” of addresses with large balances
  - Many addresses belong to exchanges and other intermediaries that hold bitcoins on behalf of many investors
- We use graph analysis and examine utilization pattern to separate intermediary and individual accounts

Intermediary Ownership



- As of Dec 2020, exchanges and other intermediaries held 5.5M BTC

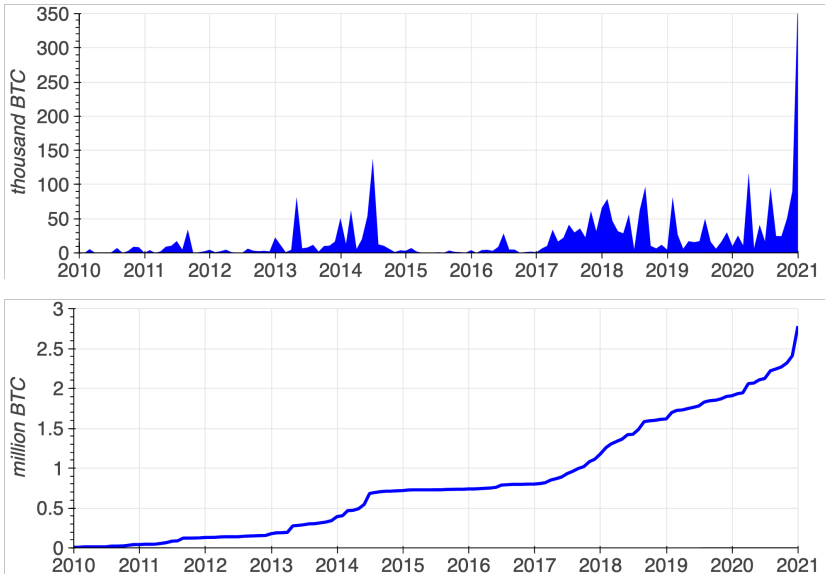
Individual Ownership



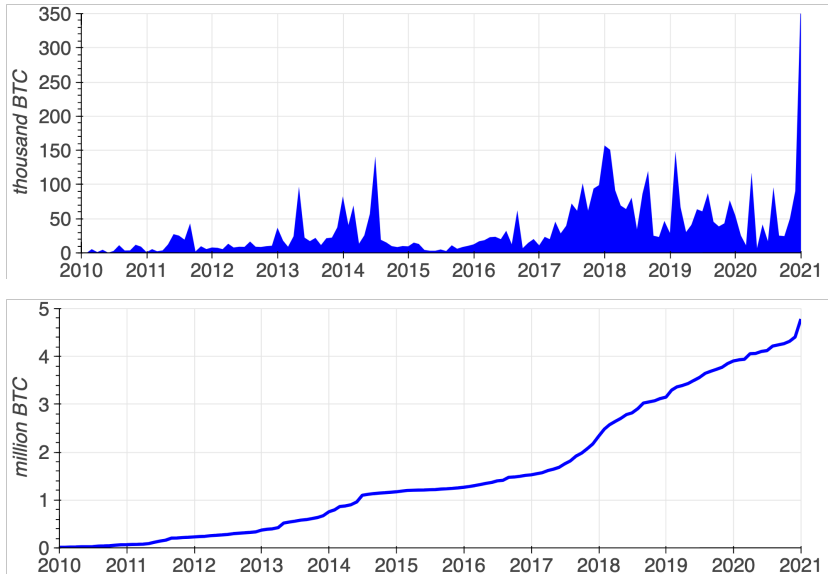
- As of Dec 2020, individuals held 8.5M BTC
- High concentration of ownership:
  - top 1000 investors control around 3M BTC
  - top 10,000 - 5M BTC

Individual Ownership: Lost coins

- Some people might have lost their private keys
- Check when an address was used last time



top 1K



top 10K

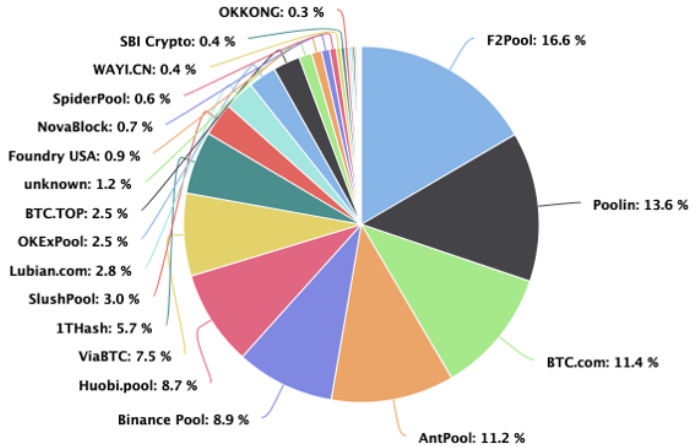
- The balance of the smallest cluster in the 1000 list is 900 BTC and in the 10,000 list is 93 BTC
- Bitcoin reached $1000 for the first time in 2014 and stayed above $1000 since 2017.

Benchmarking Concentration

- Saez and Zucman (2020) show as of 2020, the wealth share of the top 1% households in the US is more than 35% of wealth, and the top 0.1% hold about 16%.
- Estimates from Crypto.com suggest 71 million holders of Bitcoin as of January 2021. Top 7,100 holders (0.01% of all Bitcoin holders) hold 4.5 million bitcoins, which is 24% of all bitcoins in circulation
- Thus, the concentration of Bitcoin wealth is significantly higher than the wealth distribution of the US population

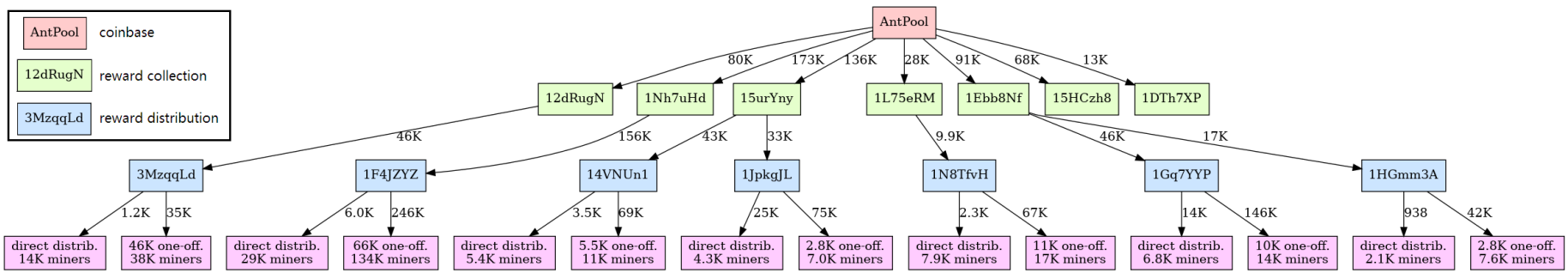
Miners

Miners: Provide Verification of Transactions

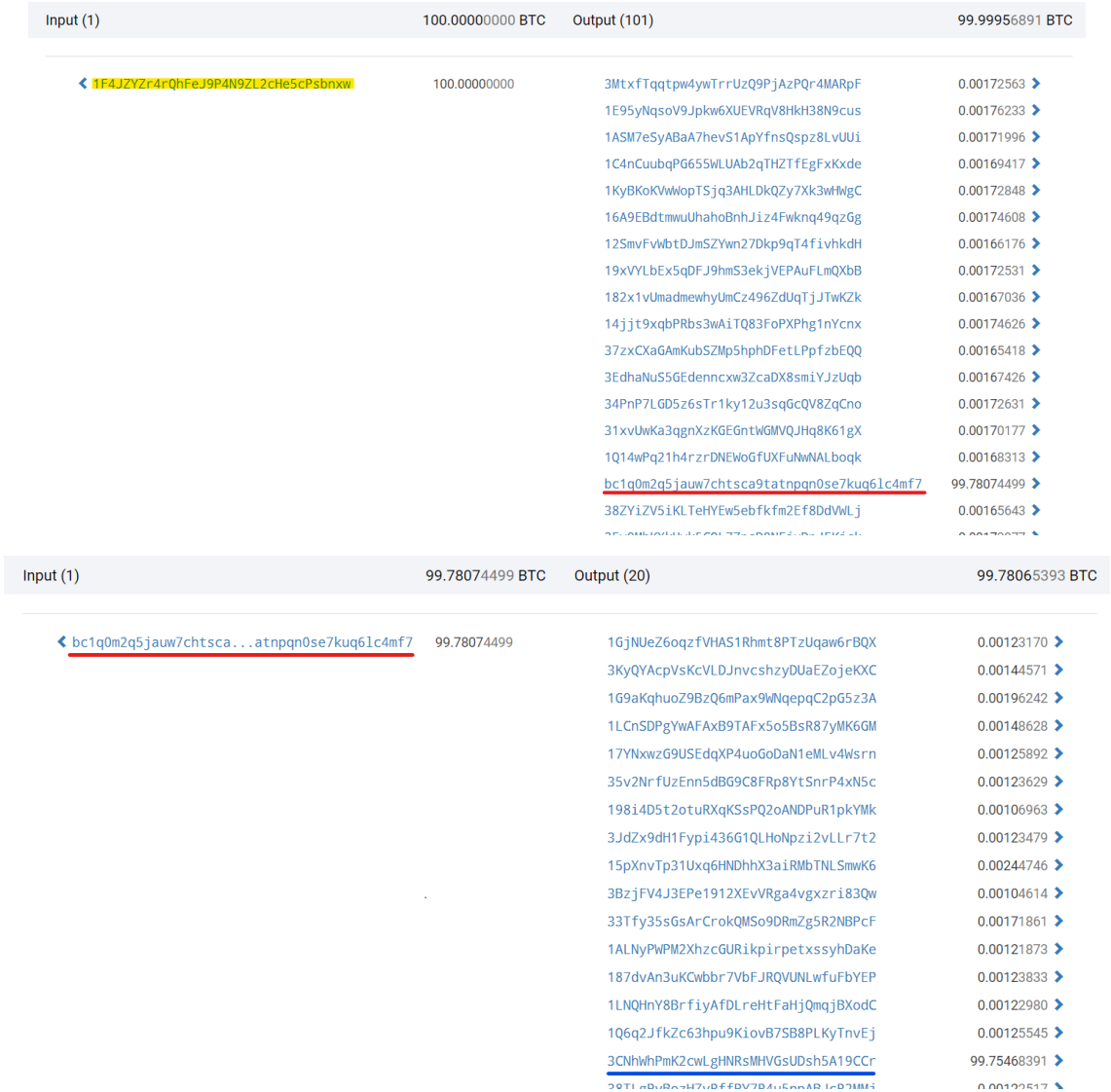


- Mining is done in pools
  - Provide coinsurance by pooling capacity of miners
  - Highly concentrated
  - Majority of pools are registered in China
- But mining pools are not miners!
- Pools' power depends on the size distribution of miners
- We identify miners by analyzing pool distributions ~ 250K miners

Mining Pool Distribution: AntPool Example



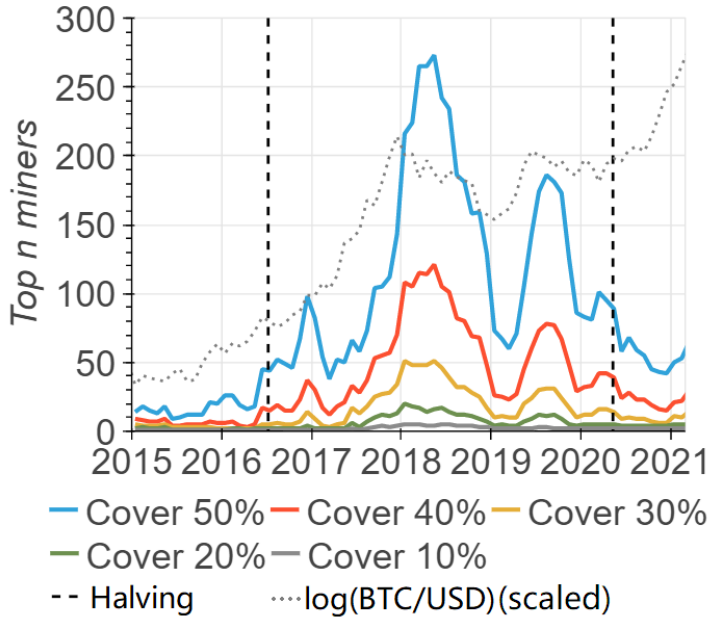
Mining Pool Distribution: AntPool Example



Input (1)	100.00000000 BTC	Output (101)	99.99956891 BTC
1F4JZYZr4rQhFeJ9P4N9ZL2cHe5cPsbnxw	100.00000000	3MtxfTqqtpw4ywTrrUzQ9PjAzPQr4MARpF	0.00172563
		1E95yNqsoV9Jpkw6XUEVRqV8HkH38N9cus	0.00176233
		1ASM7eSyABaA7hevS1ApYfnsQspz8LvUUi	0.00171996
		1C4nCuubqPG655WLUAb2qTHZTfEgFxKxde	0.00169417
		1KyBKoKVwWopTSjq3AHLDkQZy7Xk3wHWgC	0.00172848
		16A9EBdtmwuUhahoBnhJiz4Fwknq49qzGg	0.00174608
		12SmvFvWbtDJmSZYwn27Dkp9qT4fivhkdH	0.00166176
		19xVYLbEx5qDFJ9hmS3ekjVEPAuFLmQXbB	0.00172531
		182x1vUmadmewhyUmCz496ZdUqTjJTwKZk	0.00167036
		14jjt9xqbPRbs3wAiTQ83FoPXPhg1nYcnx	0.00174626
		37zxCXaGAmKubSZMp5hphDFetLPpfzbEQQ	0.00165418
		3EdhaNuS5GEdenncxw3ZcaDX8smiYJzUqb	0.00167426
		34PnP7LGD5z6sTr1ky12u3sqGcQV8ZqCno	0.00172631
		31xvUwKa3qgnXzKGEGntWGMVQJHq8K61gX	0.00170177
		1Q14wPq21h4rzrDNEWoGfUXFuNwNALboqk	0.00168313
		bc1q0m2q5jauw7chtsca9tatnpqn0se7kuq6lc4mf7	99.78074499
		38ZYiZV5iKLTeHYEw5ebfkfm2Ef8DdVWLj	0.00165643

Input (1)	99.78074499 BTC	Output (20)	99.78065393 BTC
bc1q0m2q5jauw7chtsca...atnpqn0se7kuq6lc4mf7	99.78074499	1GjNUeZ6oqzfVHAS1Rhmt8PTzUqaw6rBQX	0.00123170
		3KyQYAcpVsKcVLDJnvcshzyDUaEZojeKXC	0.00144571
		1G9aKqhuoZ9BzQ6mPax9WNqepqC2pG5z3A	0.00196242
		1LCnSDPgYwAFAxB9TAFx5o5BsR87yMK6GM	0.00148628
		17YNxwzG9USEdqXP4uoGoDaN1eMLv4Wsrn	0.00125892
		35v2NrfUzEnn5dBG9C8FRp8YtSnrP4xN5c	0.00123629
		198i4D5t2otuRXqKSsPQ2oANDPuR1pkYMk	0.00106963
		3JdZx9dH1Fypi436G1QLHoNpzi2vLLr7t2	0.00123479
		15pXnvTp31Uxq6HNDhhX3aiRMbTNLSmwK6	0.00244746
		3BzjFV4J3EPe1912XEvVRga4vgxzri83Qw	0.00104614
		33Tfy35sGsArCrokQMSo9DRmZg5R2NBPcF	0.00171861
		1ALNyPWPM2XhzcGURikpirpetxssyhDaKe	0.00121873
		187dvAn3uKCwbbr7VbFJRQVUNLwfuFbYEP	0.00123833
		1LNQHnY8BrfiyAfDLreHtFaHjQmqjBXodC	0.00122980
		1Q6q2JfkZc63hpu9KiovB7SB8PLKyTnvEj	0.00125545
		3CNhWhPmK2cwLgHNRsMHVGsUDsh5A19CCr	99.75468391

Concentration of Mining Capacity



- Mining capacity is concentrated
- Concentration varies with the Bitcoin price
- High concentration increases systemic risk and makes it easier to collude to ensure that miners continue to earn rents in equilibrium

Main Takeaways

- The majority of Bitcoin volume is for trading activity
- Exchanges are central entities on the blockchain
- The current regulation has limited effect on preventing tainted flows from entering into circulation
- The Bitcoin eco-system is still dominated by large and concentrated players:
  - Bitcoin ownership is concentrated
  - Mining industry is concentrated

Thank You!